

4-13-80

A

IBM Docket No. RAL9-00-00322

In the United States Patent and Trademark Office

Patent Application Transmittal

Transmitted herewith for filing is the Patent Application of:

Inventors(s): Brian Mitchell Bass, Teresa Lynn Davis, Anthony Matteo Gallo, Michael Steven Siegel,
Fabrice Jean Verplanken, Gail Irene Woodland

For: LOCAL MAC ADDRESS LEARNING IN LAYER 2 FRAME FORWARDING

Enclosed are

25 pages of specification, including 20 claims, plus 11 sheets of drawings.

An assignment of the invention to International Business Machines Corporation, Armonk, New York 10504.

A certified copy of a/an application.

X Declaration and Power of Attorney. NOT FULLY EXECUTED

PTO-1449 & references

X A return post card

Other:

Filing Fee Calculation (For Other Than Small Entity)

Basic Fee:						\$690.00
Claims Fees:	Filed	Limit	Extra		Rate per Extra	
Total claims:	20	20	0		\$18.00	\$0.00
Independent claims:	4	3	1		\$78.00	\$78.00
Multiple Dependent Claim Presented					\$260.00	\$0.00
Total						\$768.00

Please charge Deposit Account 09-0464 for the Total set forth above. The Commissioner is authorized to charge payment of any additional filing fees required under 37 CFR §1.16 and any patent application processing fees under 37 CFR §1.17 or to credit any overpayment to the identified account. A duplicate copy of this sheet is enclosed.

Express Mail Certificate

Express Mail Label No: EK506752165US

Date: April 11, 2000

I hereby certify that I am depositing the papers identified above with the U.S. Postal Service "Express Mail Post Office to Address" service on the above date, addressed to the Commissioner of Patents and Trademarks, Washington, DC 20231

Karen Orzechowski

Karen Orzechowski

BY:

Joselyn G. Cockburn

Joselyn G. Cockburn

Attorney of Record Reg. No. 27,069

Date: April 11, 2000

IBM Corporation 972/B656
Intellectual Property Law
PO Box 12195
Res. Tri. Park, NC 27709

Telephone: 919-543- 9036 FAX 919-543-3634

LOCAL MAC ADDRESS LEARNING IN LAYER 2 FRAME FORWARDING

**Brian M. Bass, Terri L. Davis, Anthony M. Gallo,
Michael S. Siegel, Fabrice J. Verplanken, Gail I. Woodland**

CROSS REFERENCE TO RELATED APPLICATIONS

This application is related by common inventorship and subject
matter to co-pending application titled "Method for Bridging Control
Point Data Frames Using a Network Processor". Until such time as
the foregoing application is assigned an application number by the
U.S. Patent and Trademark Office, it may be referenced by the
following applicant:attorney docket numbers: RAL9-00-0035,
421/061. The listed application is assigned to International Business
Machines Corporation and is entirely incorporated herein by this
reference.

BACKGROUND OF THE INVENTION

The present invention relates to computer networks such as
LANs (local area networks), and more particularly to a software-
based method for reducing network bandwidth consumption and
economizing on database space in MAC (media access control)
address learning.

In networks, data is typically exchanged between
communicating devices in the form of "frames." Frames include a

source MAC address and a destination MAC address; a MAC address uniquely identifies a network device in a "Layer 2" communication protocol used in Ethernet and Token Ring LANs.

A network switch connected between communicating devices uses the destination MAC address in an incoming frame generated by a source device to forward the frame to the appropriate target device. The switch typically includes a "control point" (CP) comprising a central processor and control software, and a plurality of forwarding processors. To facilitate forwarding, the CP and forwarding processors perform MAC address "learning."

MAC address learning refers to a process for reducing network traffic, wherein the source MAC address in a received frame is recorded in an address database along with its port of origin. This allows future frames destined for the source address to be forwarded only to the port on which that address is located. Otherwise, an unrecognized address must be "flooded out", i.e., forwarded to every port where the address may reside.

In existing systems, the CP maintains a central database of learned MAC addresses. A number of distributed MAC address databases which are duplicates of the central database are also maintained, locally to groups of ports and forwarding processors. As frames are received, a receiving forwarding processor performs look-ups of the source MAC addresses in the local databases, to determine whether they have been already learned. If not, the

forwarding processor notifies the CP, and in response, the CP learns the source address (i.e., records it in the central database along with its port of origin), and adds the address to every local database.

In a variation of the above, the local databases are not duplicated, but the CP is still used to learn source MAC addresses whenever a forwarding processor cannot find the address in a local database.

In another variation, the CP is not used in learning; instead, if a forwarding processor cannot find a source address in a local database, it sends control messages to all other forwarding processors, directing them to add the source address to their respective local databases.

It may be appreciated from the foregoing description that existing methods of MAC address learning entail considerable bandwidth overhead, due to the need for notifying the CP each time an unknown source address is received and for subsequently downloading the address to every local database, or for sending control messages. Also, there is unnecessary duplication of MAC address databases at distributed local sites; most local sites (i.e., physically separate groups of ports and forwarding processors) only use a small proportion of the overall MAC addresses.

In view of the foregoing, a more efficient method of MAC address learning is needed.

SUMMARY OF THE INVENTION

According to the present invention, MAC address learning is performed locally rather than by a central processor. In a preferred embodiment, a MAC address database which is local to a destination or target port for a data frame is updated by a local forwarding processor with the source address and port of origin of the data frame. Similarly, the responding data frame is learned locally, in the MAC address database corresponding to the ingress port for the initial data frame. Once the addresses are learned locally, frame forwarding between the corresponding devices can be performed more efficiently.

Local as opposed to centralized learning includes the advantages of eliminating the unnecessary duplication of a central database at local sites. Space is also conserved because each local database contains only those addresses needed and used. Processing cycles are offloaded from the CP, which no longer needs to maintain a central database or download addresses to local databases. Bandwidth that would otherwise have been needed for centralized MAC address learning is available for frame forwarding.

Additionally, MAC address "aging" may be performed locally. Aging refers to a process of deleting infrequently-used addresses from a MAC database to save space. Local aging also offloads the CP.

A method of reporting the locally-learned, distributed MAC addresses databases in an aggregate form to network users or a network management application is also disclosed.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows an example of a communication network;

Figure 2 shows components in a network switch;

Figure 3 shows a process flow in a method for local MAC address learning according to the present invention;

Figures 4A-4E illustrate the application of the method shown in Figure 3 in a specific example;

Figure 5 shows entries in a MAC database being traversed for database aging;

Figure 6 shows a process flow in aging;

Figures 7A and 7B show a process flow in a method for user-controlled MAC address reporting;

Figures 8A and 8B further illustrate MAC address reporting; and

Figure 9 shows examples of computer-usable media which may be used to store computer-executable instructions according to the present invention.

DETAILED DESCRIPTION OF THE INVENTION

Figure 1 shows an example of a typical network configuration in

which the method of the present invention would find useful application. Network devices 101 such as personal computers, servers, terminals for data entry and display, printers and the like, are connected by a physical medium 102 such as twisted pair or coax cable to nodes such as bridges 103, switches 100, routers 104, and hubs/repeaters (not shown). Data generated by end users of the network devices travels across the network to other end users via the physical medium and the nodes, which perform the processing necessary to ensure the data arrives in a usable condition with the proper recipient.

For some applications, such as LANs, networks may be confined to a relatively small geographic area, such as a building or a campus. However, large numbers of smaller networks can be linked to form extensive composite networks including thousands of devices and nodes. As devices contend for a share of the network bandwidth, data collisions can result which degrade network performance. To reduce collisions, a network may be logically partitioned into "segments", for example segments 105 and 106, connected by nodes such as bridges, switches and routers.

In general, bridges and switches transfer data frames between segments, by filtering, flooding, or forwarding the data frames. Filtering refers to dropping or discarding a received data frame when processing of the MAC addresses according to a set of rules. Flooding, as described above, refers to forwarding a data frame to all

ports where a destination address may reside, when the address has not been learned and is thus unrecognized. Forwarding refers to sending an incoming data frame having a learned MAC address to the corresponding known port or ports.

5 As noted above, MAC addresses uniquely identify a network device in a Layer 2 communication protocol. The Layer 2 protocol may be viewed as occupying a level in a "protocol stack" of established, commonly-used communication standards, beginning with a Layer 1, representing the physical medium used to transfer data, followed by Layer 2, then by a Layer 3 and 4 and so on. Generally speaking, the layers above Layer 2 handle exchanges between network entities where the exchanges require more processing than at the lower levels.

10 Figure 2 shows a switch 100 in greater detail. The switch includes a CP 200 and a plurality (not limited to three) of cards 202-204, connected to the CP 200 and to each other via a pipe 201 over which frames and control messages flow and which may be implemented as a DASL crossbar switch or a bus. Each card includes a local MAC database 208, a plurality (not limited to four) of forwarding processors 207, a specialized processor 209 known as a Guided Tree Handler (GTH), and a plurality of ports 210. Each card further includes a storage medium such as a ROM or RAM 205 for storing computer-executable instructions, referred to herein as "picocode" 206, which when executed by a forwarding processor and

GTH, implement frame forwarding and MAC address learning. The GTH also executes the picocode to perform address aging and reporting functions detailed hereinafter.

The switch 100 is connected by a physical medium 102 between network segments 105 and 106. When a network device 101 needs to communicate with another device in a different network segment, data is transferred between the devices, via the physical medium, in frames such as frame 211. To enable the frame to arrive at the correct destination as it travels across the network, header information is added to the front of the frame in fields corresponding to the various protocols used (L1 = Layer 1, L2 = Layer 2, etc.). The header information is followed by the actual user data and a trailer field.

According to the present invention, when a frame 211 enters a receiving port 210, a forwarding processor 207 on the card executes picocode 206 causing the method illustrated in flowchart form in Figure 3 to be implemented. Referring now to Figure 3, the forwarding processor for the receiving port performs a look-up in the local MAC address database 208 for the destination address, as shown in blocks 300 and 301. As shown in blocks 302-304, if the address is found, the frame is forwarded via the connecting bus 201 to the card containing the target port or ports where the destination address is located. If not, the frame is flooded out to all possible target ports.

When a card containing a target port receives a frame forwarded to it from another card on the switch (block 305), local MAC address learning at the destination (i.e., learning on the target card) according to the present invention is performed as shown in
5 blocks 305-312. As shown in blocks 306 and 307, a local processor 207 performs a look-up in the local MAC address database 208 for the source address. If the source address is not found, the processor notifies the GTH 209 that there is a new address to be learned, or, if the port of origin of the source address has changed,
10 the processor notifies the GTH 209 that the address information must be updated.

As noted above, the GTH is a specialized processor for accessing and maintaining the local MAC address database. More specifically, MAC addresses may be stored in a database structure
15 known as a "Patricia tree" which enables fast searching and access. The tree entries are also chained together in a singly-linked circular list. While any processor may perform a search of the MAC address database, the GTH is used to add, delete or update database entries, or perform any actions which require following the entry chain. (The
20 GTH and the local processor can communicate without using any of the bandwidth used for forwarding frames.) Upon being notified by a local processor as described above, the GTH records the source MAC address of the forwarded frame in the local database, along with its port of origin, as shown in block 308.

20

As shown in Figure 4C, the source address is not present in the local MAC database on card 204, so the source address and port of

origin are added to the database as entry 401, and the frame is forwarded out port y to addr2. (It is noted that the representation of entries in a MAC database 208 has been simplified in the Figures for purposes of illustration.)

5 As shown in Figure 4D, typically the target device sends a frame in response. In the responding frame 402, the MAC addresses are reversed. What was formerly the source address is now the destination address, and what was formerly the destination address is now the source address. Similarly, the roles of the receiving and target ports are now reversed. Accordingly, when a local processor for the receiving port (port y) looks up the destination address for the responding frame, the local MAC database of the receiving port contains the destination address (addr1) and target port (port x), since it was just added by the process described above. The responding frame may thus be forwarded directly to the target port.

10 As shown in Figure 4E, when the card containing the target port for the responding frame receives the forwarded frame, it performs the method shown in Figure 3, learning the source address (addr2) and port of origin (port y) as described above, as shown in entry 403. Thus, subsequent exchanges between the network devices corresponding to addr1 and addr2 may be performed more efficiently, since the addresses have been learned at the corresponding local sites on the switch.

In view of the foregoing, the advantages of learning MAC

addresses locally at a target port, as opposed to using a CP, are evident. Unnecessary duplication of databases is avoided, since each local database contains only those addresses needed and used. Further, bandwidth that would otherwise have been needed for MAC address learning using a CP is available for frame forwarding.

An additional benefit of local MAC address learning is realized in an "aging" process which occurs for MAC addresses. Because, as noted above, networks can be very extensive, MAC address databases can become very large. Accordingly, in the aging process, database entries corresponding to addresses which have not been accessed for a pre-determined period of time are purged in order to free up space.

According to the present invention, aging is performed locally. As shown in Figure 5, the switch hardware records an access to an address when a look-up of a source MAC address is performed on a MAC database 208 and an entry is found, by setting an "entry seen" bit in the entry. An aging task resets the seen bit periodically. An entry is considered to be expired if the entry seen bit has not been set since the last time the entry was visited by the aging task. Each GTH traverses the local database to delete expired entries.

A process flow for aging is illustrated in Figure 6. The MAC database entries are circularly linked in the order they are learned. Because MAC address databases can be extremely large, the GTH

database and eliminates the records of any accesses. At the end of the accelerated aging period, the GTH deletes any entry which does not have a record of an access during the accelerated aging period.

Topology change notifications are common events which can result in purging of entire databases, causing extreme stress to switches that perform aging from the CP. Thus, it can be seen that performing aging locally significantly relieves the CP.

The present invention further provides a method of controllably reporting on the status of the local MAC address databases to users, if desired. The method compiles the distributed MAC databases into an aggregate database, which may be the required form for reporting to network management applications on the switch. It may also be used for console display for viewing by humans.

To provide for address reporting as needed, computer-executable instructions according to the present invention perform a "mirroring" method. Figures 7A and 7B show a process flow for the method, with Figure 7A showing steps executed by the CP, and Figure 7B showing steps executed by a GTH.

If an optional mirroring function is enabled, the CP will start a report timer having a timer value configurable by a user. After a period of time defined by the timer value elapses, the CP will send a Begin Report Request to each GTH, as shown in block 700. As shown in blocks 710 and 711 of Figure 7B, upon receipt of the request, each GTH bundles a pre-defined number of MAC addresses

into a Begin Report Reply and sends the bundle to the CP, recording the last address sent (this address cannot be deleted while it is recorded as the last address). To locate a starting address for the Begin Report Reply, a GTH will perform a search on a default MAC address which is always present in the database, for instance, the BPDU address (the BPDU address is defined by the 802.1D standard and is used by the Spanning Tree algorithm). This address acts as a logical beginning of a chain of MAC addresses stored in the database.

As shown in blocks 701-704 in Figure 7A, when the CP receives the Begin Report Reply, it will add each MAC address into a mirror database if the address is not already in the database. Each MAC address in the mirror database has a timer value associated with it. When an address is first entered, the value is positive. Every time the MAC address is reported again the timer value will be reset to the positive value.

After a subsequent period determined by the report timer, the CP will send a Continue Report Request to each GTH (block 705). As shown in blocks 712 and 713, when a GTH receives this request, it obtains the MAC address chained after the last MAC address it reported, bundles that address and a pre-determined number of the addresses following it into a Continue Report Reply which it sends to the CP. The GTH records the last address sent.

As shown in blocks 706 and 707, as the CP receives the

the GTHs form bundles 803 of a pre-determined number of MAC addresses and send them to the CP, recording a last address sent 804. The CP compiles an aggregate mirror database 802 from the bundles, which may be reported out to a user on a display screen, for example, or reported to a network manager.

As noted above, the mirroring function is optional and configurable by a user. A user can set the reporting time interval in accordance with the level of traffic on the switch so as to achieve a tolerable level of reporting interference.

Figure 9 illustrates examples of computer-executable media such as diskette 900, CD-ROM 901, magnetic tape 902 or hard disk 903 which may be used to store and transport computer-executable instructions according to the present invention. As is well understood, the instructions may be retrieved from the storage media and executed by a processor to effect the method of the invention. Typically, a forwarding processor 207 or GTH 209 executes the picocode 206. The CP 200 typically executes a high-level language such as "C" code 904.

The foregoing description of the invention illustrates and describes the present invention. Additionally, the disclosure shows and describes only the preferred embodiments of the invention, but it is to be understood that the invention is capable of use in various other combinations, modifications, and environments and is capable of changes or modifications within the scope of the inventive concept as

What is Claimed Is:

1 1. A method comprising:
2 receiving a data frame on a receive port from a first
3 device connected to a network, said data frame including a source
4 media access control (MAC) address for said first device and a
5 destination MAC address for a second device connected to said
6 network;
7 forwarding said data frame to a target port corresponding
8 to said second device; and
9 learning said source MAC address locally to said target
10 port.

1 2. The method of claim 1, said learning step comprising:
2 determining whether said source MAC address is present
3 in a database local to said target port; and
4 recording said source MAC address in said local
5 database if not present.

1 3. The method of claim 1, further comprising performing
2 frame forwarding using said locally-learned MAC address.

1 4. The method of claim 2, further comprising aging said
2 MAC address locally.

1 6. The method of claim 5, further comprising:
2 compiling a plurality of said replies into an aggregate database;
3 and
4 reporting said aggregate database to a network user or
5 manager.

1 7. The method claim 5, wherein said report request is
2 issued at time intervals which are configurable by a network user.

1 8. A network switch comprising:
2 a port connected to a network device;
3 processors and a MAC address database local to said
4 port;
5 said processors programmed to perform MAC address
6 learning locally to said port.

9. A network switch comprising:
a port connected to a network device;
processors and a MAC address database local to said port;

6 media access control (MAC) address for said first device and a
 7 destination MAC address for a second device connected to said
 8 network;
 9 forwarding said data frame to a target port corresponding
 10 to said second device; and
 11 learning said source MAC address locally to said target
 12 port.

1 15. The computer-usable medium of claim 14, said learning
 2 step comprising:
 3 determining whether said source MAC address is present
 4 in a database local to said target port; and
 5 recording said source MAC address in said local
 6 database if not present.

1 16. The computer-usable medium of claim 14, said method
 2 further comprising performing frame forwarding using said locally-
 3 learned MAC address.

1 17. The computer-usable medium of claim 14, said method
 2 further comprising aging said MAC address locally.

1 18. The computer-usable medium of claim 14, said method
 2 further comprising:

3 bundling a pre-determined number of said MAC
4 addresses into a reply in response to a report request from a control
5 point; and
6 transmitting said reply to said control point.

1 19. The computer-usable medium of claim 18, said method
2 further comprising:

3 compiling a plurality of said replies into an aggregate
4 database; and

5 reporting said aggregate database to a network user or
6 manager.

1 20. The computer-usable medium of claim 18, wherein said
2 report request is issued at time intervals which are configurable by a
3 network user.

LOCAL MAC ADDRESS LEARNING IN LAYER 2 FRAME FORWARDING

ABSTRACT OF THE INVENTION

5

The present invention relates to a software-based method for reducing network bandwidth consumption and economizing on database space in MAC (media access control) address "learning." In networks, data is typically exchanged between communicating devices in the form of "frames." Frames include a source MAC address and a destination MAC address; a MAC address uniquely identifies a network device in a "Layer 2" communication protocol used in Ethernet and Token Ring LANs. In MAC address learning, the source MAC address in a received frame is recorded in an address database along with its port of origin to facilitate subsequent frame forwarding. In typical existing systems, a central "control point" (CP) processor learns and maintains a central database of MAC addresses. A number of distributed MAC address databases which are duplicates of the central database are also maintained. This centralized approach entails considerable bandwidth overhead and unnecessary duplication of MAC address databases at distributed local sites. According to the present invention, MAC address learning is performed locally rather than by a central processor. In a preferred embodiment, a MAC address database which is local to a

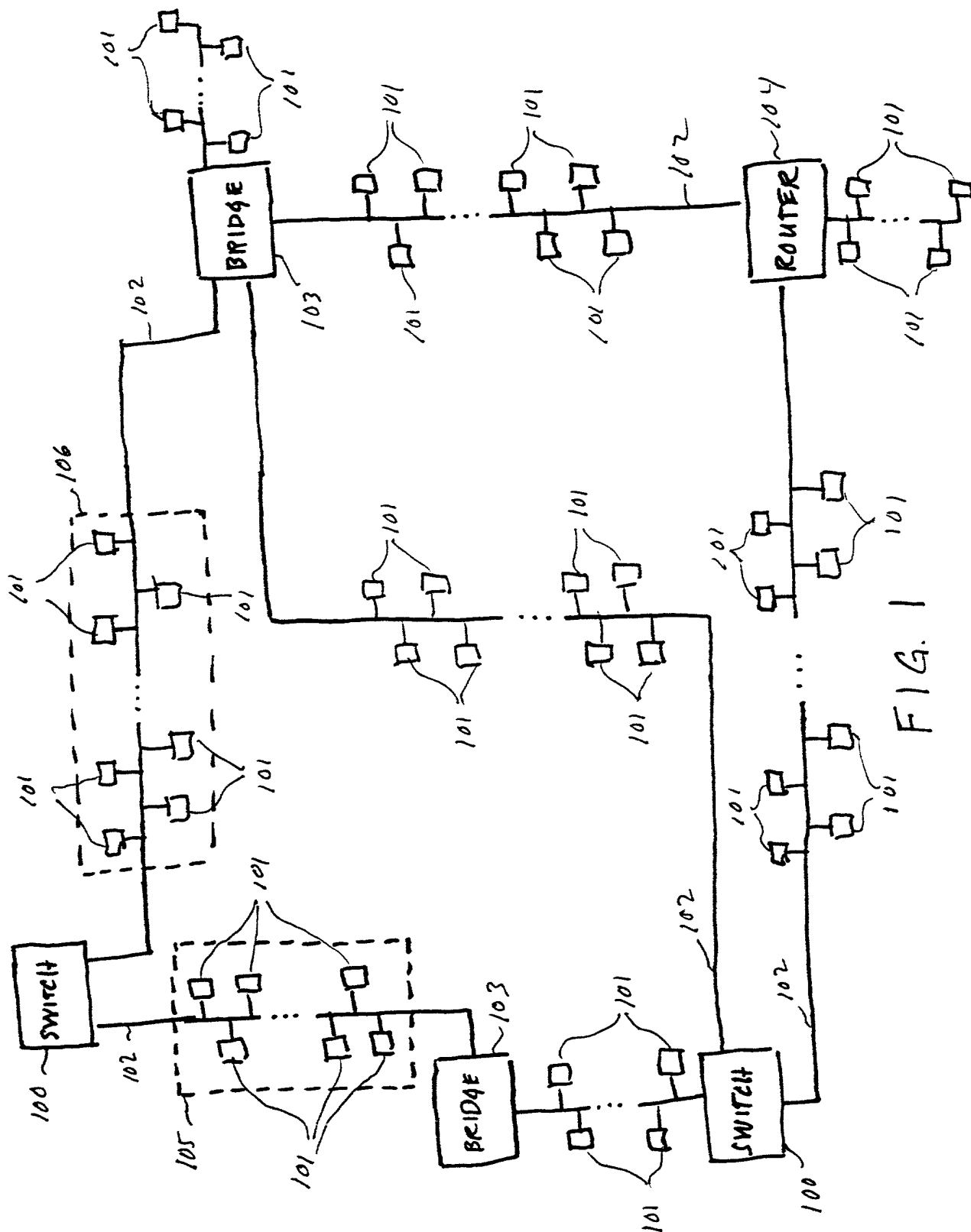
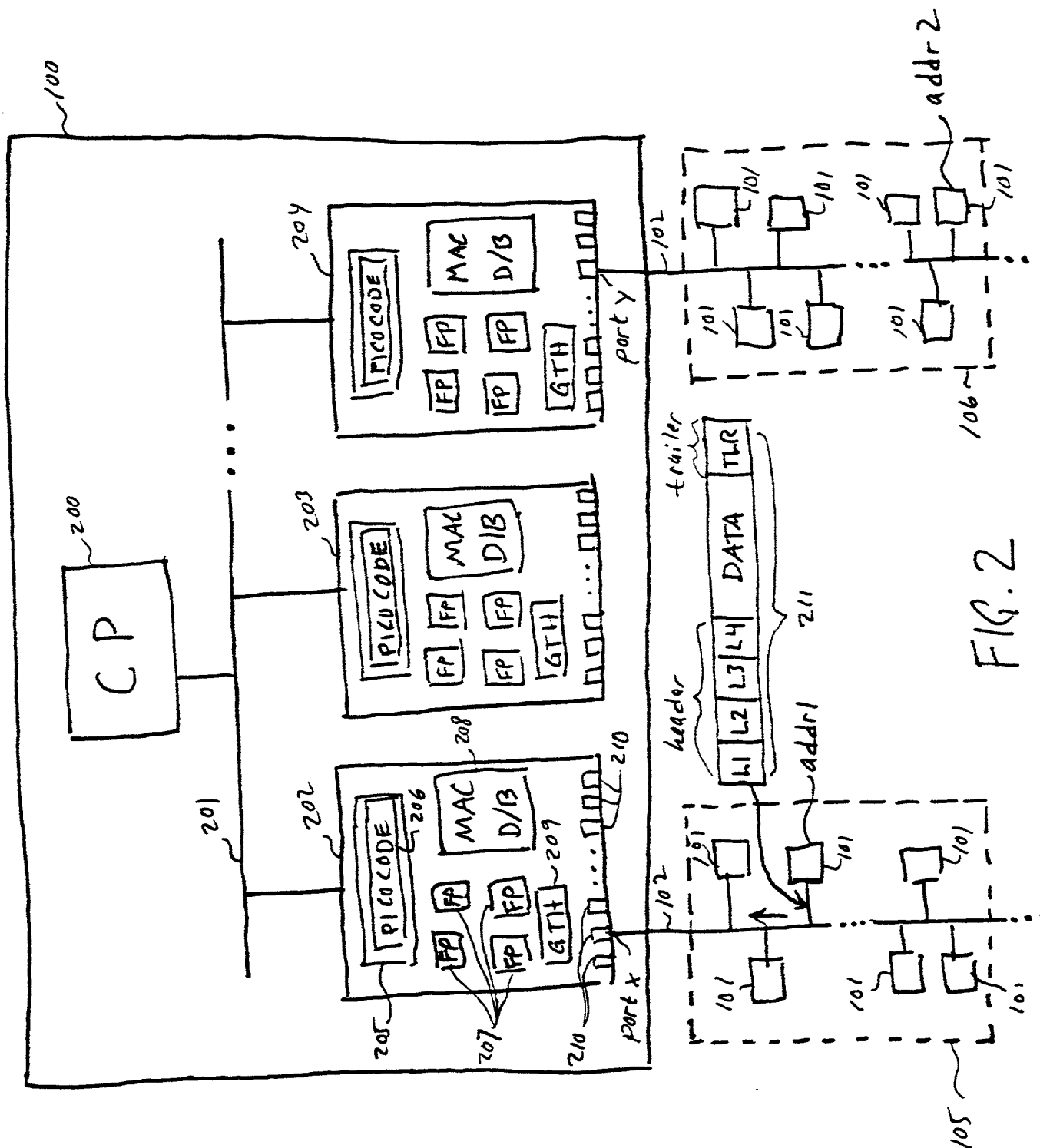
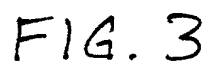


FIG. 1





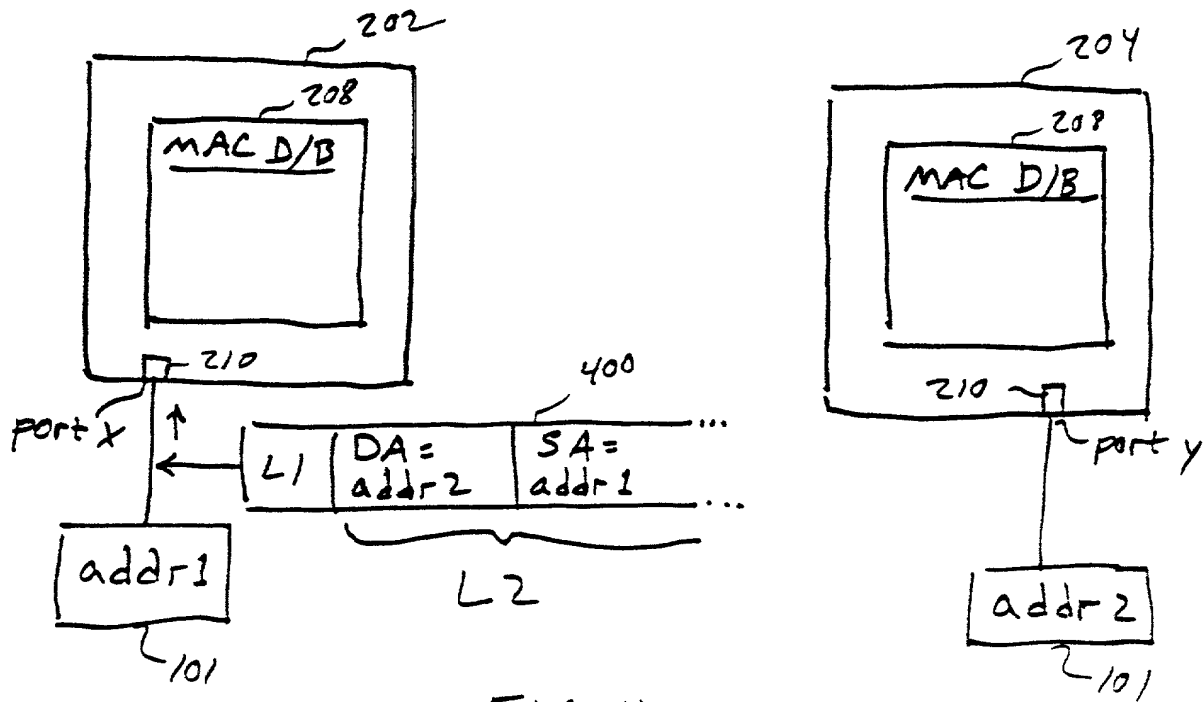


FIG. 4A

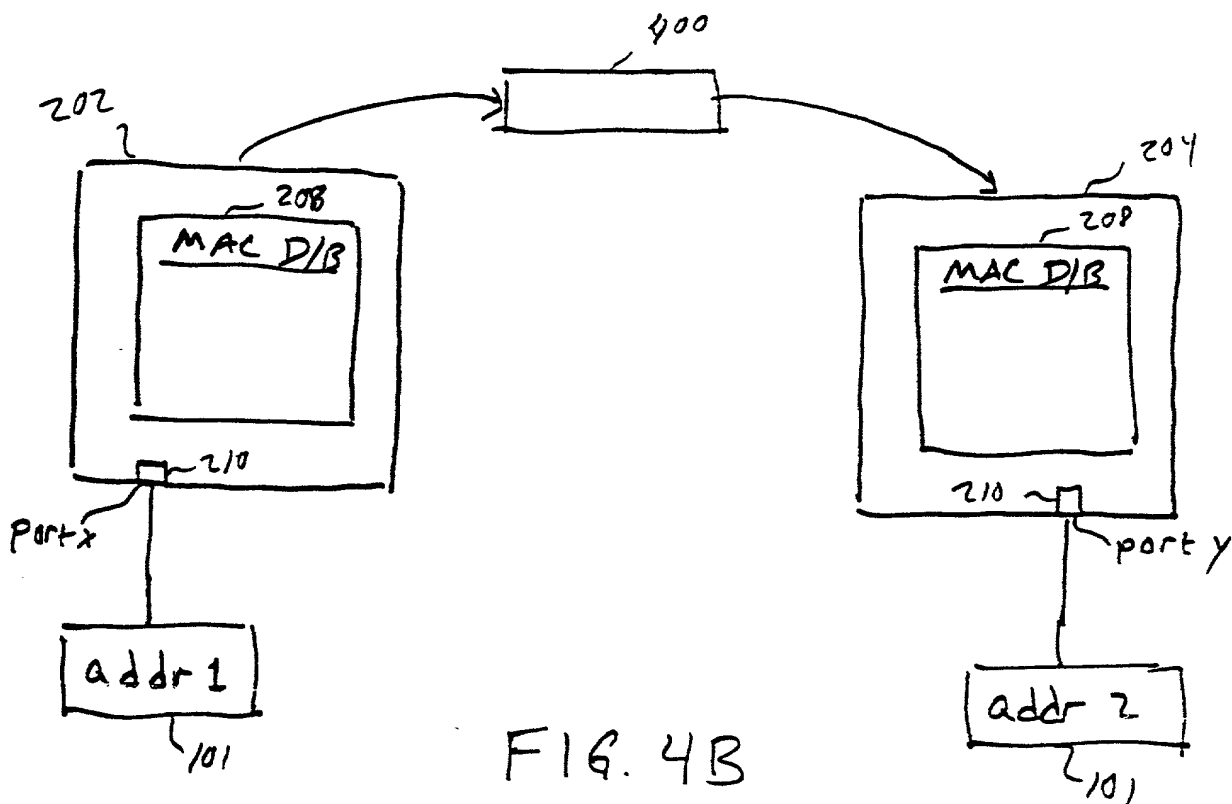
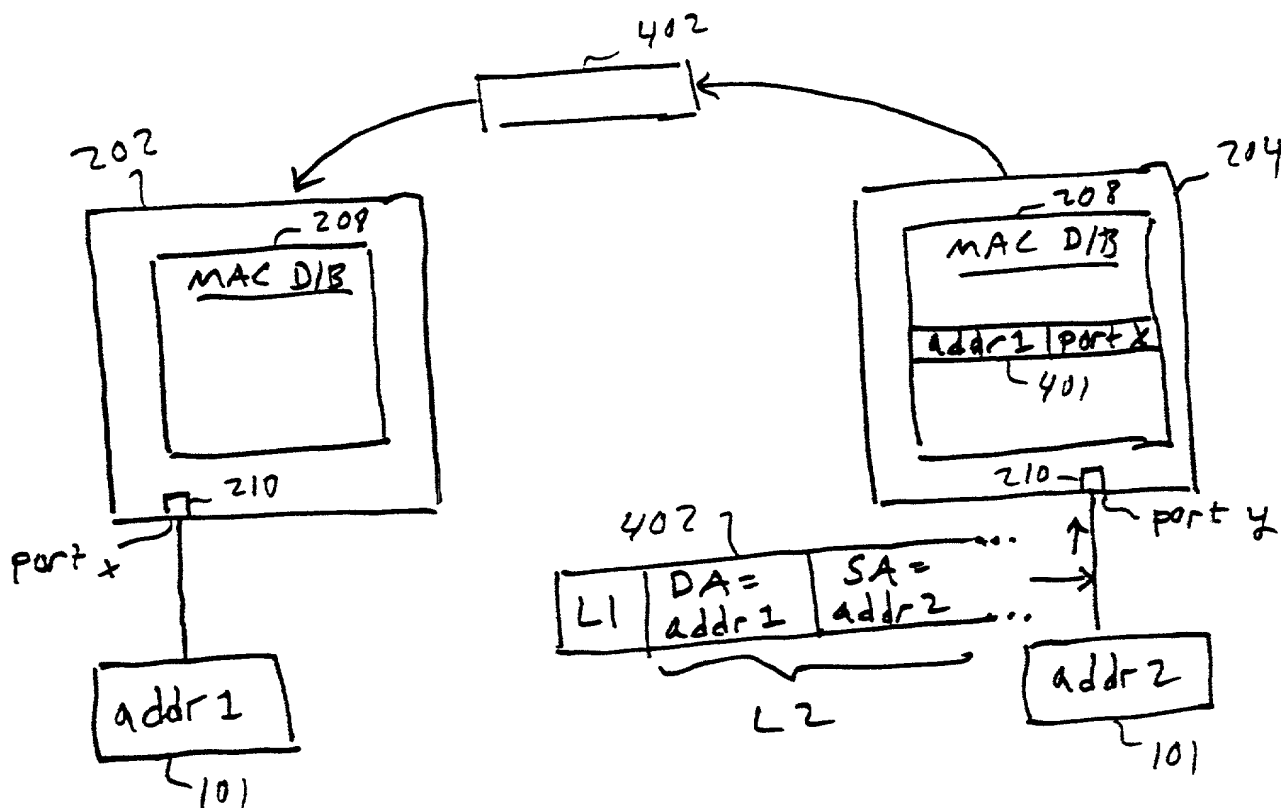
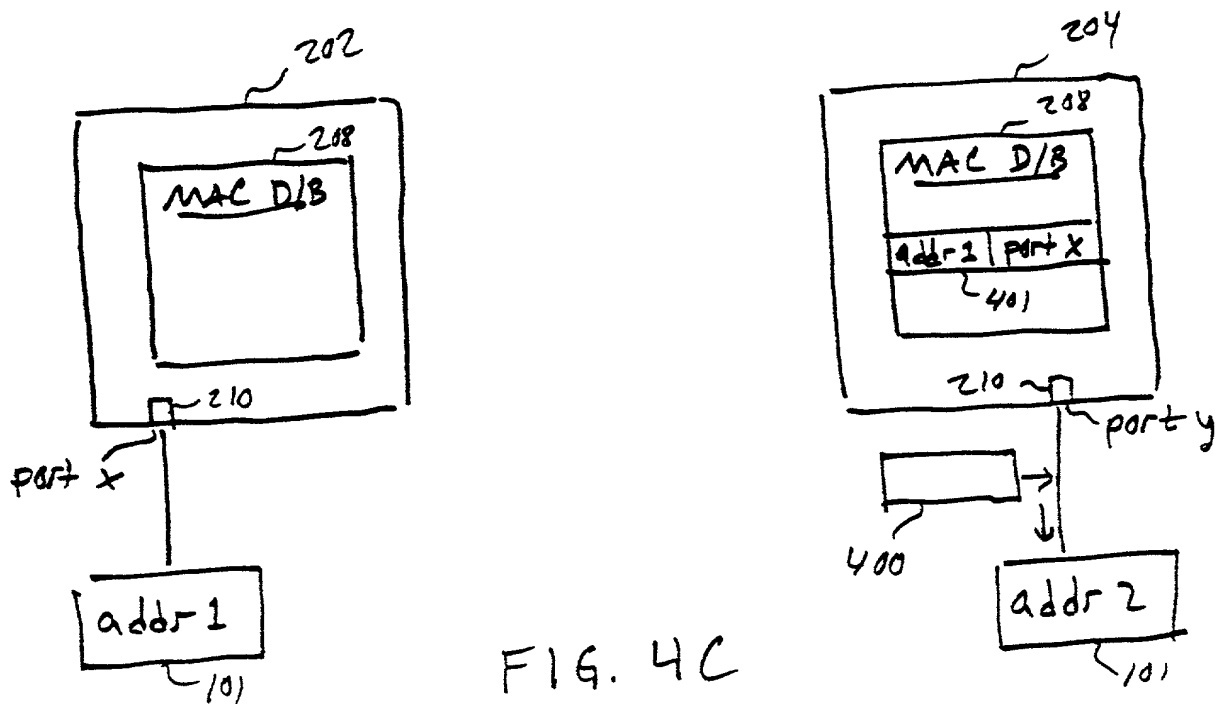


FIG. 4B



09547369 044100

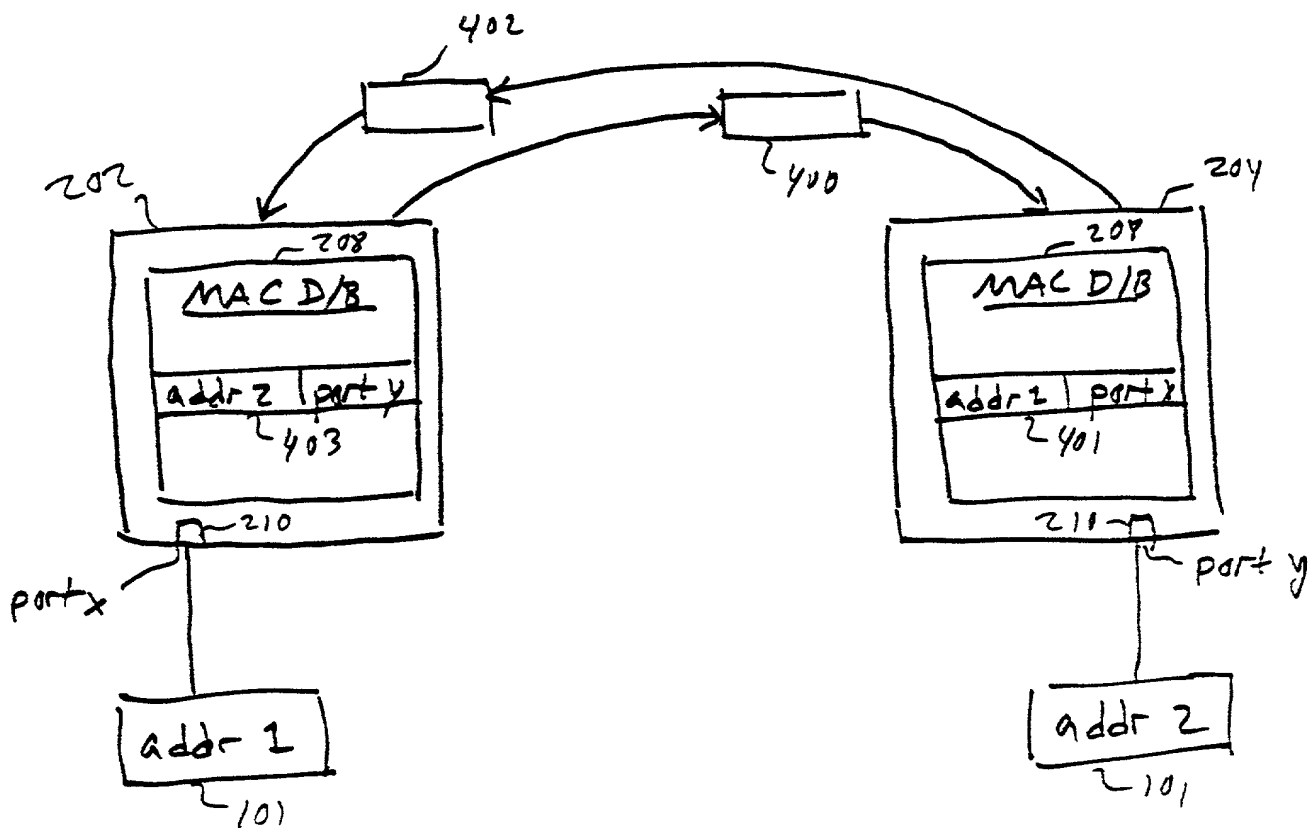


FIG. 4E

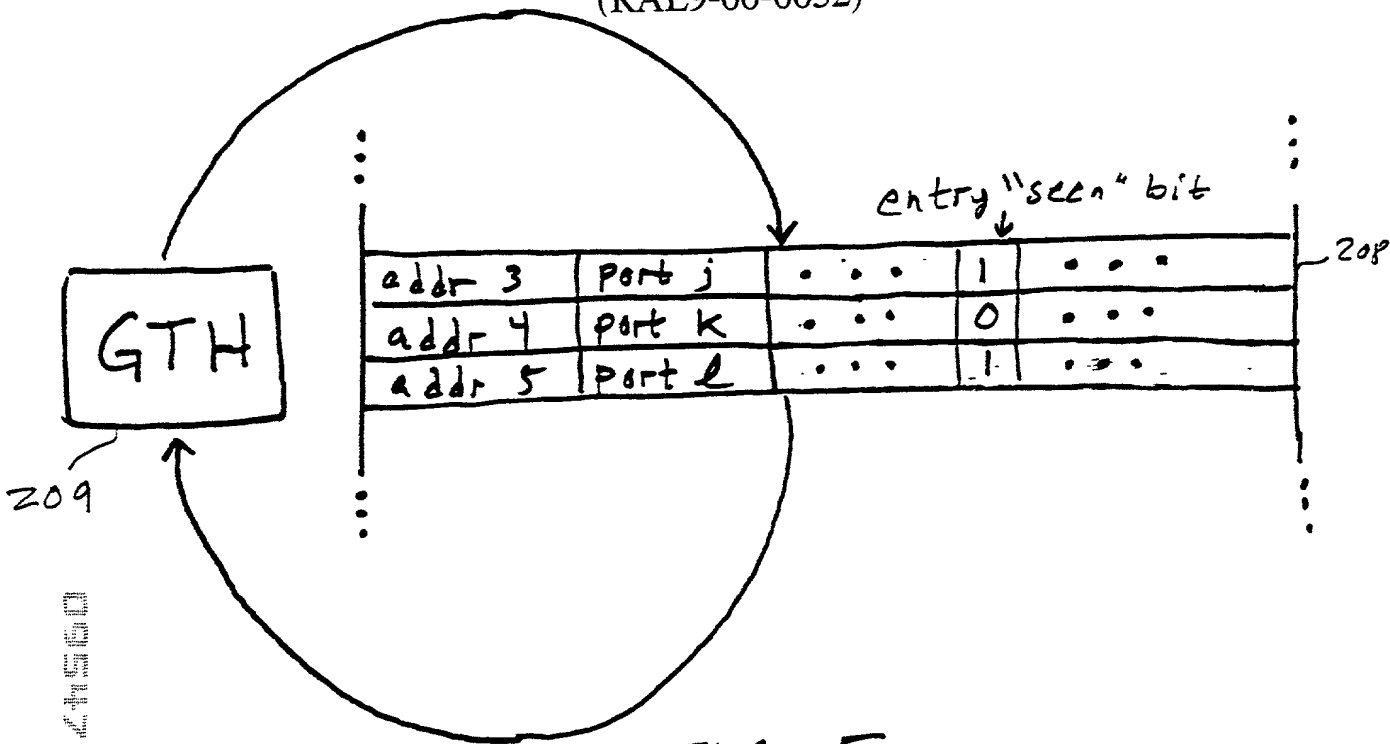


FIG. 5

(While there are more entries to examine:)

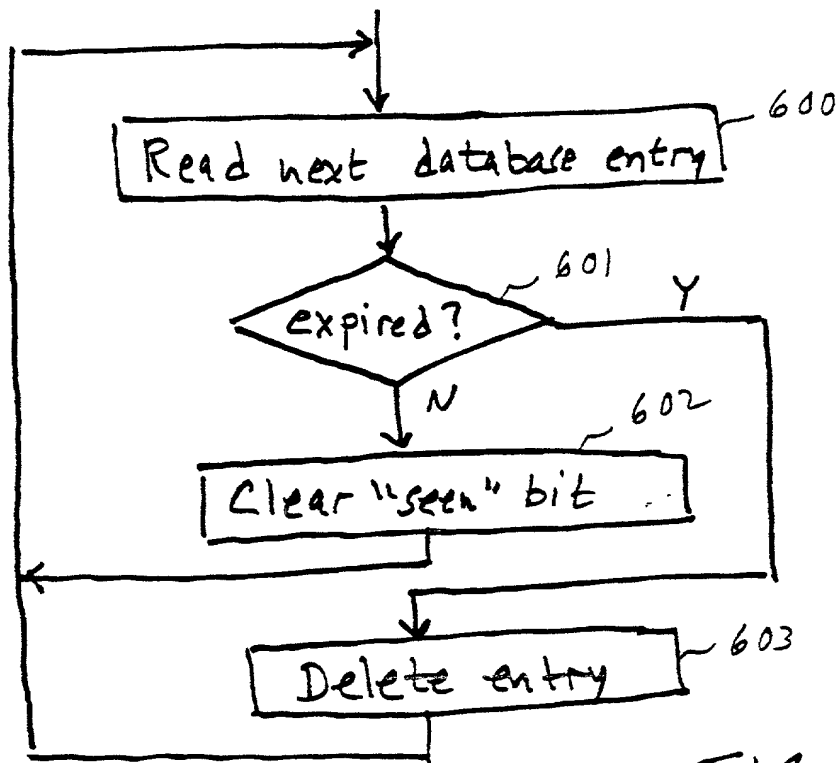


FIG. 6

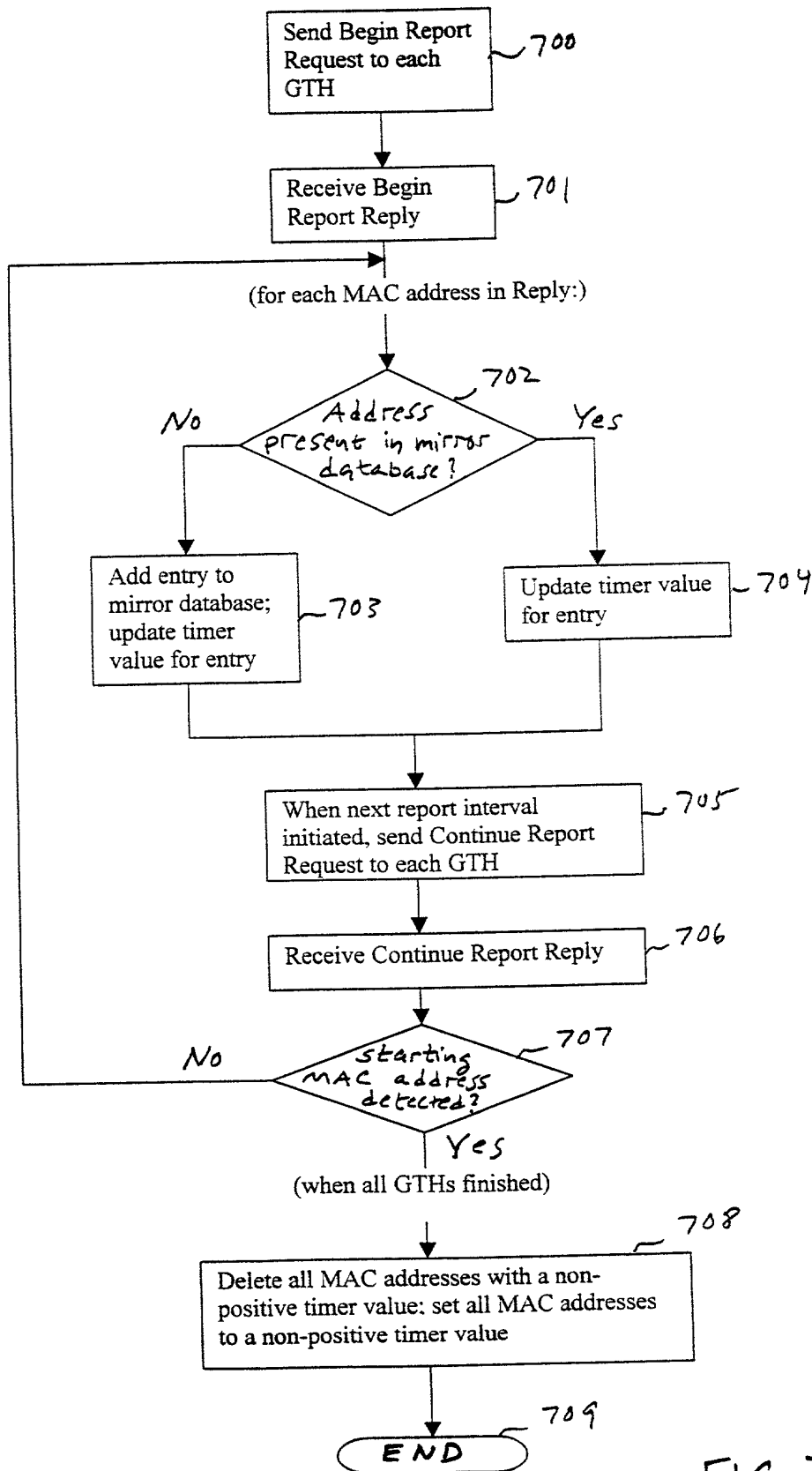


FIG. 7A

9/11
(RAL9-00-0032)

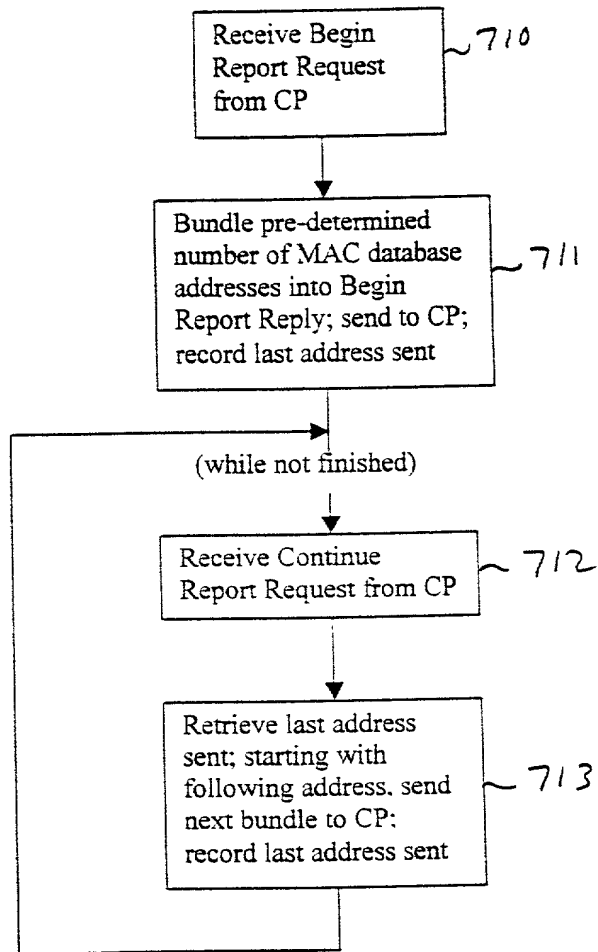


FIG. 7B

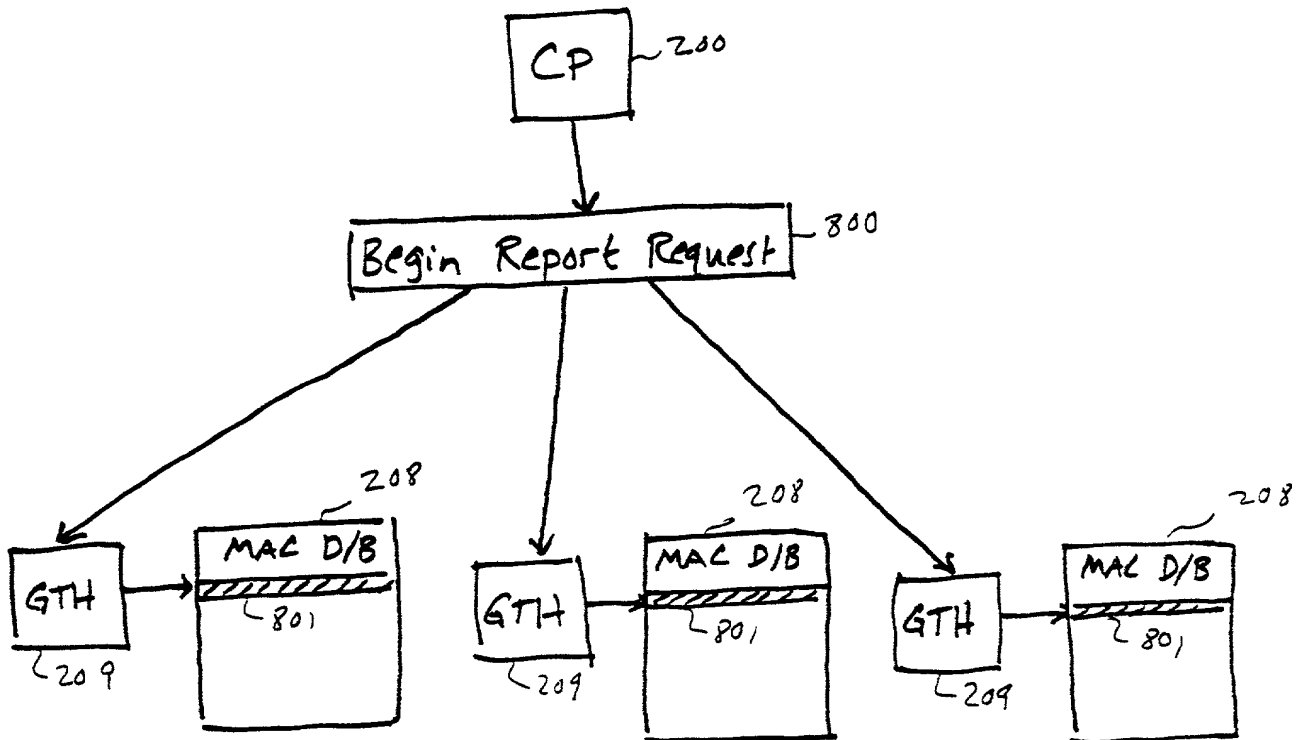


FIG. 8A

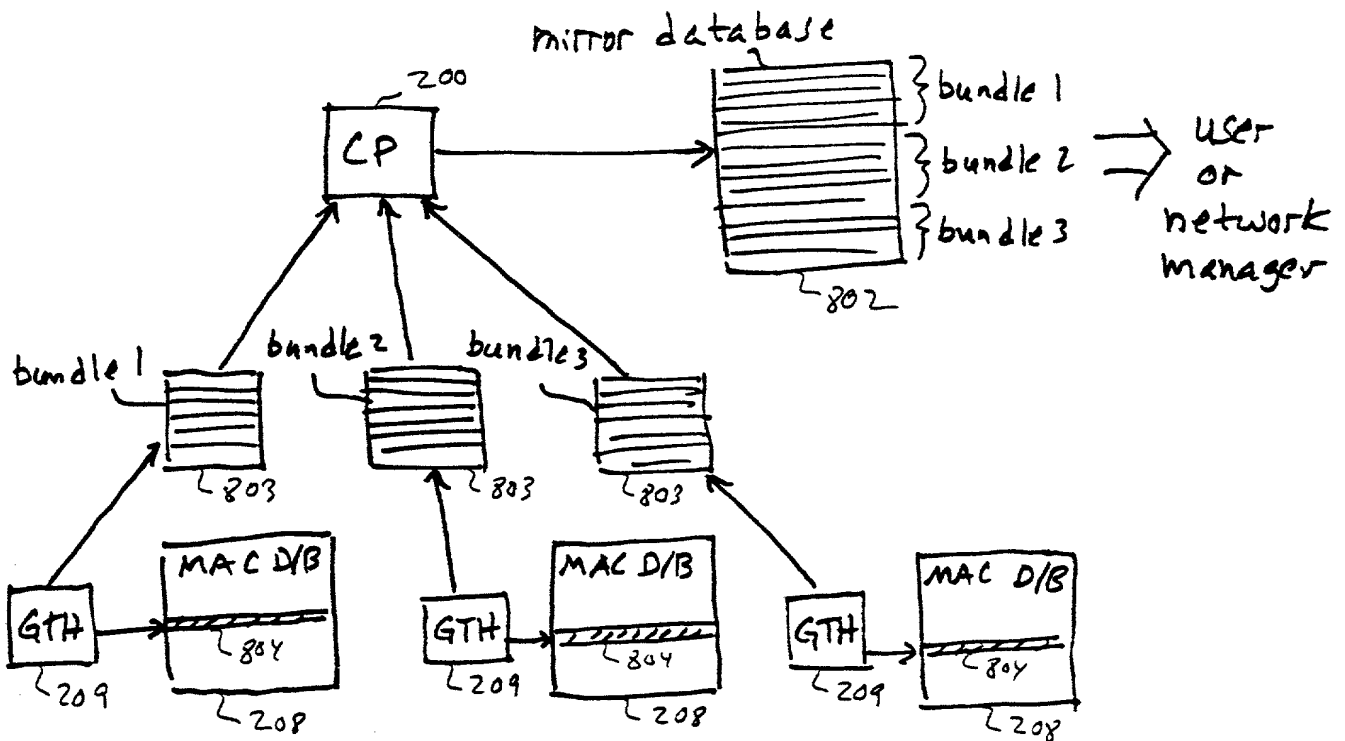


FIG. 8B

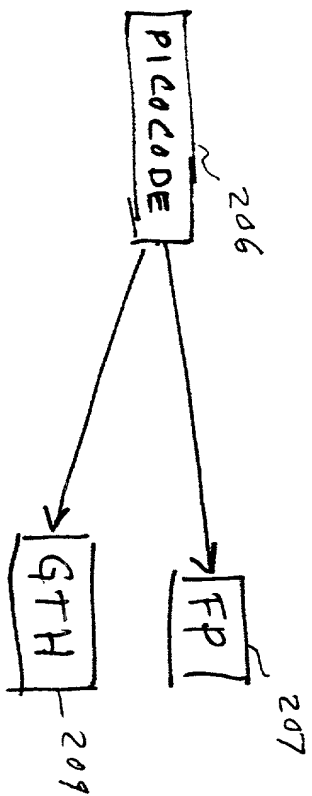
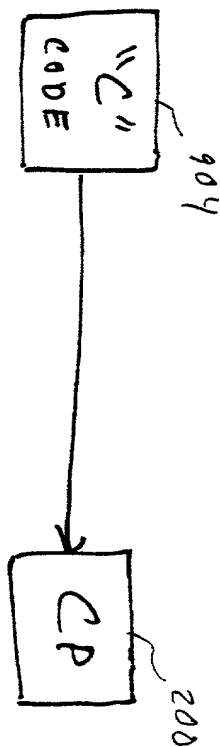
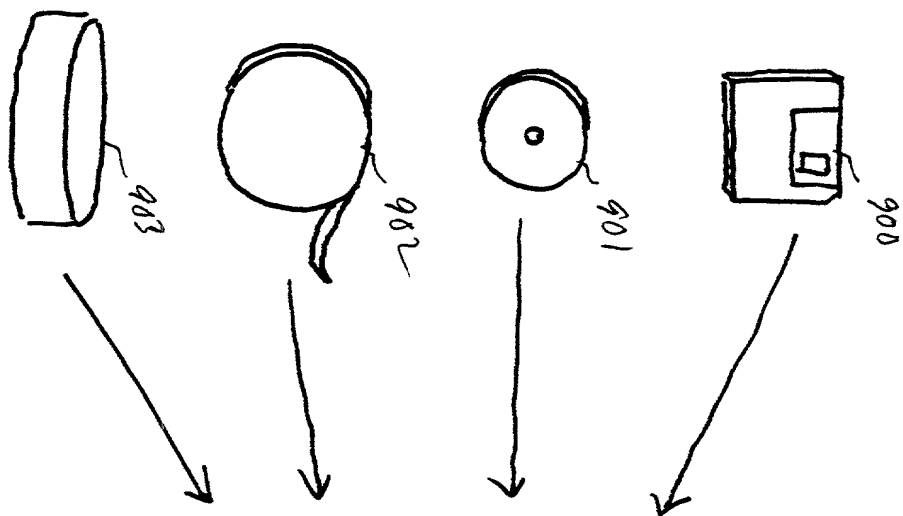


FIG. 9

**DECLARATION AND POWER OF ATTORNEY
FOR PATENT APPLICATION**

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name; I believe I am an original, first and joint inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled:

LOCAL MAC ADDRESS LEARNING IN LAYER 2 FRAME FORWARDING

the specification of which is identified by the attorney (IBM) Docket Number appearing above.

I hereby state that I have reviewed and understand the contents of the above- identified specification, including the claims.

I acknowledge the duty to disclose information which is material to the patentability of this application in accordance with Title 37, Code of Federal Regulations, §1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

Prior Foreign Application(s)

<u>Number</u>	<u>Country</u>	<u>Day/Month/Year</u>	<u>Priority Claimed</u>
---------------	----------------	-----------------------	-------------------------

I hereby claim the benefit (a) under Title 35, United States Code, §119(e) of any U.S. application listed below and identified as a provisional application or (b) under Title 35, United States Code, §120 of any U.S. application listed below and not identified as a provisional application, and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior U.S. application in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose information material to the patentability of this application as defined in Title 37, Code of Federal Regulations, §1.56 which occurred between the filing date of the prior application and the national or PCT international filing date of this application

Prior U.S. Applications

<u>Serial No.</u>	<u>Filing Date</u>	<u>Status</u>
-------------------	--------------------	---------------

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Send all correspondence to: Joscelyn G. Cockburn, IBM Corporation 972/B656; PO Box 12195; Research Triangle Park, NC 27709.

First Inventor: Brian Mitchell Bass

Signature: _____ **Date** _____

Residence: 4021 Old Sturbridge Drive, Apex, NC 27502

Citizenship: USA

Post Office Address: same as above

Second Inventor: Teresa Lynn Davis

Signature: _____ *Date* _____

Residence: 8217 Canthy Circle, Garner, NC 27529

Citizenship: USA

Post Office Address: same as above

Third Inventor: Anthony Matteo Gallo

Signature: _____ *Date* _____

Residence: 3308 Corsham Drive, Apex, NC 27502

Citizenship: USA

Post Office Address: same as above

Fourth Inventor: Michael Steven Siegel

Signature: _____
Date

Residence: 10625 Lowery Drive, Raleigh, NC 27615

Citizenship: USA

Post Office Address: same as above

Fifth Inventor: Fabrice Jean Verplanken

Signature: _____
Date

Residence: 9152 Route de Cagnes, 06610 La GAUDE, France

Citizenship: France

Post Office Address: same as above

Sixth Inventor: Gail Irene Woodland

Signature: _____
Date

Residence: 123 Booth Road, Chapel Hill, NC 27516

Citizenship: USA

Post Office Address: same as above